

Diagnostic reasoning prompts reveal the potential for large language model interpretability in medicine

Year: 2023 | Citations: 201 | Authors: Thomas Savage, Ashwin Nayak, Roberta Gallo

Abstract

One of the major barriers to using large language models (LLMs) in medicine is the perception they use uninterpretable methods to make clinical decisions that are inherently different from the cognitive processes of clinicians. In this manuscript we develop diagnostic reasoning prompts to study whether LLMs can imitate clinical reasoning while accurately forming a diagnosis. We find that GPT-4 can be prompted to mimic the common clinical reasoning processes of clinicians without sacrificing diagnostic accuracy. This is significant because an LLM that can imitate clinical reasoning to provide an interpretable rationale offers physicians a means to evaluate whether an LLMs response is likely correct and can be trusted for patient care. Prompting methods that use diagnostic reasoning have the potential to mitigate the “black box” limitations of LLMs, bringing them one step closer to safe and effective use in medicine.